High-resolution structural insights on the sugar-recognition and fusion tag properties of a versatile β-trefoil lectin domain from the mushroom *Laetiporus sulphureus*

Iván Angulo[1,8], Iván Acebrón[1], Blanca de las Rivas[2], Rosario Muñoz[2], I. Rodríguez-Crespo[3], Margarita Menéndez[4,5], Pedro García[5,6], Hiroaki Tateno[7,9], Irwin J. Goldstein[7], Begoña Pérez-Agote[1], and José M. Mancheño[1]*

[1]Departamento de Cristalografía y Biología Estructural. Instituto de Química Física Rocasolano. CSIC; Serrano 119, E-28006 Madrid. Spain

[2]Grupo de Biotecnología Bacteriana, Instituto de Ciencia y Tecnología de Alimentos, CSIC, Juan de la Cierva 3, E-28006 Madrid. Spain

[3]Departamento de Bioquímica y Biología Molecular, Facultad de Químicas, Universidad Complutense de Madrid, E-28040 Madrid. Spain

[4]Grupo de Macromoléculas. Instituto de Química Física Rocasolano. CSIC; Serrano 119, E-28006 Madrid. Spain

[5]Centro de Investigación Biomédica en Red de Enfermedades Respiratorias, Bunyola, Mallorca, Illes Balears, Spain

[6]Departamento de Microbiología Molecular y Biología de las Infecciones. Centro de Investigaciones Biológicas. CSIC; Ramiro de Maeztu 9, E-28040 Madrid. Spain

[7]Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, Michigan 48109-0606

[8]Present address: The Scripps Research Institute, Scripps Florida, Jupiter, FL-33458, USA

[9]Present address: Research Center for Medical Glycoscience, Advanced Industrial Science and Technology (AIST), Ibaraki 305-8568, Japan

*To whom correspondence should be addressed: Tel: +34-91-7459547; Fax: +34-915642431; e-mail: xjosemi@iqfr.csic.es

Running title: Structural and functional study of the lectin module LSL_{150}

Keywords: crystal structure/ lactose/ lectin/ solubility tag/ water rearrangement

Abstract

In this work, we analyzed at high-resolution the sugar-binding mode of the recombinant N-terminal ricin-B domain of the haemolytic protein LSLa (LSL_{150}) from the mushroom *Laetiporus sulphureus*, and also provide functional *in vitro* evidences suggesting that, together with its putative receptor-binding role, this module may also increase the solubility of its membrane pore-forming partner. We firstly demonstrate that recombinant LSL_{150} behaves as an autonomous folding unit and an active lectin. We have determined its crystal structure at 1.47 Å resolution, and also that of the [LSL_{150}:(lactose)β, γ)] binary complex at 1.67 Å resolution. This complex reveals two lactose molecules bound to the beta and gamma sites of LSL_{150}, respectively. Isothermal titration calorimetry indicates that LSL_{150} binds two lactoses in solution with highly different affinities. Also, we test the working hypothesis that LSL_{150} exhibits *in vivo* properties typical of solubility tags. With this aim, we have fused an engineered version of LSL_{150} (LSL_t) to the N-terminal end of various recombinant proteins. All the designed LSL_{150}–tagged fusion proteins were successfully produced at high yield and, furthermore, the target proteins were purified by a straightforward affinity procedure on agarose-based matrices due to the excellent properties of LSL_{150} as affinity tag. An optimized protocol for target protein purification was devised which involved removal of the LSL_{150} tag through in-column cleavage of the fusion proteins with His_6-tagged TEV endoprotease. These results permitted to set up a novel, lectin-based system for production and purification of recombinant proteins in *E. coli* cells with attractive biotechnological applications.

Introduction

The crystal structure of the haemolytic lectin LSLa from the mushroom *Laetiporus sulphurous* (Mancheño et al. 2005) revealed a homohexameric assembly composed of protein subunits (~35 kDa) with a modular architecture: an N-terminal (residues 1-150) β-trefoil lectin module (Nt-LSL_a) and a C-terminal (residues 151-314) membrane pore-forming module (PFM). This last module showed three dimensional similarities to the PFM of the aerolysin-like pore-forming toxins revealing LSLa as a β pore-forming toxin, namely, it forms oligomeric transmembrane β-barrels in their target membranes (Song et al. 1996; Melton et al. 2004) Conversely, the N-terminal lectin domain shows a β-trefoil fold, a highly conserved architecture observed in other lectins (Hazes 1996) and in toxins that bind glycoproteins, such as ricin (Rutenber et al. 1987), abrin (Tahirov et al. 1995), the hemagglutinin component (HA1) of the progenitor toxin from *Clostridium botulinum* (Inoue et al. 2003), and the pore-forming toxin CEL-III from *Cucumaria echinata* (Uchida et al. 2004). These structural/functional relationships support a direct role in cell receptor-binding for the lectin module of LSLa, and therefore the characterization of the fine details of the sugar-recognition process is required for a thorough understanding of the biological mechanism of action of this haemolytic lectin. In fact, as shown for legume lectins, ligand binding is governed by a variety of finely tuned effects (Elgavish and Shaanan 1998). In particular, since the interactions of carbohydrate with residues at the combining site are either direct or mediated by water molecules (Weis and Drickamer 1996), a high-resolution structural analysis of unliganded and liganded lectin forms should provide reliable insights on the mechanism of ligand binding, specially on subtle protein conformational changes and rearrangement of water molecules within the binding sites (Rini et al. 1993; Svensson et al. 2002; Nurisso et al. 2010). In fact, water arrangements have been identified as an important contributor to the thermodynamics of ligand binding to lectins (Chervenak and Toone 1994; Toone 1994).

On the other hand, a distinct feature of the LSLa structure is that few interactions have been identified between the N-terminal lectin module and its C-terminal pore-forming module (Mancheño et al. 2005). In contrast, the crystal structures of proaerolysin (Parker et al. 1994) from *A. hydrophila*, parasporin (Akiba et al. 2006) from *Bacillus thuringiensis*, and ε-toxin (Cole et al. 2004) from

Costridium perfringens revealed that the structural elements equivalent to those of the PFMs from LSLa are interrupted by sequence stretches which are intertwined with N-terminal segments (Mancheño et al. 2010). Despite these structural and functional features of LSLa suggest that the lectin and pore-forming modules are strictly independent structural domains, the finding that N-terminal domains of multidomain proteins may function *in vivo* as solubility enhancers for their C-terminal partners (Kim et al. 2007), and also that solubility tags such as MBP and NusA act as passive partners in the folding of target proteins (Nallamsetty and Waugh 2006), led us to consider a similar scenario for LSLa.

Here, we demonstrate that LSL_{150} is an autonomous folding unit and an active lectin when recombinantly produced in *Escherichia coli* cells, and investigated the binding mode of lactose to LSL_{150} at high resolution. Novel rearrangements of water molecules upon sugar binding have been identified. In addition, an indirect *in vitro* approach to analyze the potential of LSL_{150} as solubility tag by fusing LSL_{150} to various proteins. These proteins have been successfully produced and purified. Finally, the excellent properties of LSL_{150} as an affinity tag on Sepharose® matrices, have permitted to design a straightforward, generic and cost-effective purification protocol for recombinant target proteins in *E. coli* cells.

Results and Discussion

LSL_{150} is an autonomous folding unit

A first requirement for the validity of our working hypothesis considering the lectin module of LSLa as a solubility tag for its accompanying membrane-interacting module is that it must behave as an independent folding unit. Although it is assumed that folding principles identified from single-domain proteins are generally applicable to isolated protein modules (Han et al. 2007), herein we demonstrate the "autonomous folding unit" character of LSL_{150} by determining the crystal structure at 1.47 Å resolution of the protein recombinantly produced in *E. coli* cells. One LSL_{150} molecule is in the asymmetric unit (Figure 1). The final refined model contains 149 residues and 252 water molecules and has an *R*-factor of 15.7% and an *R*-free of 19.4%. The average value of the interface areas between crystallography-related monomers calculated with the PISA server (Krissinel and Henrick 2007) is in the range of 129-396 $Å^2$ (below 5% of the overall monomer solvent-accessible area), indicating that monomeric species are present in the crystal. This agrees with gel-filtration chromatography results on 10/30 HR™ Superdex 75 (GE Healthcare) which revealed that LSL_{150} behaves in solution as a species of ~17 kDa and with the crystal structure of LSLa, showing that the hexameric assembly was maintained by interactions between C-terminal domains (Mancheño et al. 2005).

The structure of LSL_{150} is virtually identical to that reported for Nt-LSL_a from *L. sulphureus* (Mancheño et al. 2005); both structures superpose with Cα r.m.s. deviation of 1.17 Å for 147 pairs of structurally equivalent residues (Supplementary data, Figure S1). Hence, and as previously described for Nt-LSLa, LSL_{150} assumes a β-trefoil fold displaying the characteristic pseudo-3-fold symmetry arising from tandem α, β, and γ peptide repeats (Murzin et al. 1992; Hazes 1996). The three subdomains form a six-stranded anti-parallel β-barrel capped at one end by three hairpin turns called the hairpin triplet (Figure 1). Interestingly, the structure of LSL_{150} reported here corresponds to the first unliganded structure of this β-trefoil since the previously reported for native LSLa has a lactose molecule bound to the γ-site (Mancheño et al. 2005). Therefore, LSL_{150} offers for the first time the

opportunity to evaluate the existence of conformational changes triggered by sugar binding (see below).

An updated structural comparison of LSL_{150} with known folds using the DALI algorithm (Holm and Sander 1993) reveals a high similarity to other β-trefoils occurring in a diverse set of proteins (overall rmsd found was 2.5 Å for 127 common Cα atoms for 73 unique hits with Z-score > 10) that generally share little or no detectable sequence similarity (average sequence identity 11.5%). These two aspects of β-trefoils have been interpreted in terms of convergent evolution (Loris 2002), which is evidenced by the fact that the sugar-binding sites differ significantly (see below). In fact, although the β-trefoil fold is often termed a galactose binding fold (Hirabayashi et al. 1998), it can display binding sites with very different specificities (Notenboom et al. 2002).

All these crystallographic results reveal that LSL_{150} is an autonomous folding unit when produced in *E. coli* BL21 (DE3) cells. In this sense we have estimated that the average yield of LSL_{150} production is ~80-100 mg per one-litre culture, which compares well with previous results (Tateno and Goldstein 2003) (60 mg per litre of culture) obtained with *E. coli* Nova Blue (DE3) cells for a deletion mutant of LSLa containing the first 187 amino acids (LSL_{187}). Remarkably, this deletion mutant bound Sepharose® 4B along the purification process, revealing its sugar-binding properties. This indicates that the lectin module is properly folded, in agreement with our structural results. As a whole, these results indicate that the nucleotide sequence coding for the N-terminal lectin module of LSLa is a highly translated one, and also that the cellular environment provided by *E. coli* cells is highly competent for its efficient and proper folding. These properties are key features of solubility tags (Malhotra 2009) which supports our working hypothesis that the N-terminal ricin-B domain of LSLa may function as a solubility enhancer *in vivo*.

Lactose binding in solution

As can be deduced from the crystal structure of $LSL_{150,}$ this protein binds lactose in solution, as has been shown by isothermal titration calorimetry (ITC). ITC experiments were carried out at 25 ºC in 20 mM Tris-HCl, pH 8.0, 100 mM NaCl and 0.04% (*w/v*) sodium azide. A typical binding profile is

shown in Figure 2. Lactose binding is exothermic, as indicated by the downward peaks, and the final region of the titration curve clearly deviated from the behavior expected for the single-site binding model, suggesting the presence of additional site/s with very low affinity for lactose. The analysis of the binding isotherm with the two-set of sites model yielded K_{d1} = 170 ±10 μM (ΔG_1 = -5.11 ± 0.04 kcal/mol) and ΔH_1 = 12.0 ± 0.4 kcal/mol and a lactose:LSL_{150} stoichiometry of 0.93 ± 0.02 for the higher affinity site, while the low saturation fraction of the low-affinity sites reached in the ITC experiments just allowed a rough estimation of K_{d2} ~ 11 mM (ΔG_1 ~ -3 kcal/mol) and ΔH_2 ~ -8 kcal/mol fixing to 1 the binding stoichiometry. These results agree well with previous studies reported for LSLa (Mancheño et al. 2005), showing a relative high affinity γ-site (in terms of site occupancy) and a relative low affinity β-site.

The thermodynamic parameters show that lactose binding to LSL_{150} is enthalpically driven (ΔS_1 = -6.9 ± 0.4 kcal/(mol·K) and ΔS_2 ~ -5 kcal/(mol·K)). The large negative changes in the enthalpy result from the formation of favourable protein-lactose hydrogen bonds and van der Waals interactions (see below), which would compensate for the negative change in entropy due to the loss of translational and conformational entropy of the ligand or the protein side chains at the binding sites (Dam and Brewer 2002).

High-resolution X-ray analysis of lactose binding

The binding of lactose was further characterized by the crystal structure of the ([LSL_{150}:(lactose)β, γ]) complex. Crystals of the lactose ternary complex were obtained by cocrystallization in the same mother liquor as unliganded LSL_{150}, supplemented with 0.2 M lactose. The structure of the complex was solved by molecular replacement at a resolution of 1.67 Å and refined to an *R*-factor of 17.6% and an *R*-free of 22.9%. The final refined model contains one LSL_{150} molecule, two lactose molecules, one glycerol and 262 water molecules. The quality of the resulting electron density map was excellent (Figures 3A and 3B). The total atomic B-factors of the atoms of the sugars and their liganding residues are quite similar. Thus, sugar sites appear to be fully occupied. The high resolution of the diffraction data permits the interpretation of many details in the electron density maps, including alternative

conformations for protein side-chains or hydroxyl groups of the carbohydrate. Both α and β anomers of the Glc O1 atom could be modeled and refined, with respective occupancies of 0.40 and 0.60 for the two lactose molecules of the binary complex. Occupancy values were adjusted manually to give the best results, as judged by the cleanliness of $F_o - F_c$ difference maps around the O1 atom. In solution, the α/β ratio for glucose is 0.38/0.62 (Vyas et al. 1994), thus suggesting that in the crystal state LSL_{150} shows no preferential interactions for a given anomer. Indeed, the O1 atom of the glucose unit of the lactose at the β site (in both the α and the β anomers) is hydrogen-bonded with protein atoms of a symmetry related protein molecule [α anomer: the Oγ atom of Ser-29 (2.9 Å); β anomer: the NH2 atom of Arg-90 (2.8 Å)] and in addition, a water-mediated interaction is also observed with the Oε-1 atom of Asp-56 from another symmetry related molecule (Figure 3C). Interestingly, the analysis of the electron density of Ser-29 indicates that this side chain adopts a double conformation with similar occupancies to those of the glucose anomers and, therefore, this behavior is probably related to the interaction with the O1 atom of the α anomer. On the other hand, a similar pattern of interactions is observed between LSL_{150} and the O1 atom of the glucose subunit of the lactose at the γ-site for both anomers and LSL_{150} (Figure 3D). In this case, the O1 atom of the glucose unit of the lactose (in both the α and the β anomers) is hydrogen-bonded with two solvent molecules (2. 8 Å in both cases), which in turn interact with other water molecules or with the NH1 atom from Arg-76 (2.8 Å), respectively (Figure 3D).

Additional interactions observed between the glucose unit and LSL_{150} involve the oxygen atoms O2, O3, and O6. Thus, in both the β and γ binding sites, the glucose oxygen atom O2 interacts with LSL_{150} through bridging water molecules, namely with the NH2 atom of Arg-75 (β site) and with the Nε1 atom of Trp-131 (γ site) (Figures 3C and 3D). Besides, the glucose oxygen atom O3 at the β site interacts with the same water molecule than the O2 atom, and forms a hydrogen bond with the NH1 atom of Arg-75 (3.1 Å), and also participates in a water-bridged contact with the Gln-84 main chain oxygen (Figure 3C). In turn, the glucose oxygen atom O3 at the γ site directly interacts with the NH1 and NH2 atoms of Arg-123 (2.9 and 2.8 Å, respectively), and also with a tightly bound water molecule (2.6 Å) that makes hydrogen bonds with the Asn-132 main chain oxygen (2.7 Å), and with

the Nδ1 atom of His-133 (2.7 Å), both acting as acceptor partners, and also with the NH2 atom of Arg-76 (2.7 Å) which acts as donor (Figures 3D). Interestingly, the comparison between the unliganded and liganded states of LSL_{150} revealed a concerted reorganization of a cluster of up to three water molecules at the γ site (see below) as a result of ligand binding, which is not observed at the β site. This reorganization most likely contributes to ligand binding and therefore to the different affinity exhibited by these two binding sites, as it has been reported previously for other lectins (Chervenak and Toone 1994). Finally, the O6 atom of the glucose unit bound to the γ site participates in a water-bridged contact with the Oε1 atom of Glu-137, which in turn interacts with the Nε2 atom of His-133 (2.9 Å). These contacts probably contribute to maintaining the side chain of this this last residue in the proper orientation for hydrogen bonding to the above mentioned critical water molecule.

The nonreducing galactose unit is buried in both binding sites and participates in hydrophobic stacking interactions with the aromatic rings of Tyr-91 (β site) and Phe-139 (γ site), respectively (Figures 3E and 3F). In both cases, the apolar patches formed by the 3, 4, 5, and 6 carbons of the galactose units pack against the above side chains. In addition, oxygen atoms O4 and O6 stabilize the position of the sugar in both sites: first, the oxygen atom O4 within the β site makes hydrogen bonds with the NH1 atom of Arg-75 (3.0 Å), with the Oδ2 atom of Asp-93 (2.7 Å), and with an ordered water molecule (2.9 Å); secondly, the oxygen atom O6 creates hydrogen bonds with the Oδ1 atom of Asp-93 (2.6 Å) and with the Oδ1 atom of Asn-94 (2.8 Å) (Figure 3E). On the other hand, the axial O4 oxygen of the galactose unit at the γ site is at hydrogen bond distance to the Oδ2 atom of Asp-141 (2.5 Å), to the NH1 atom of Arg-123 (2.9 Å), and to the Nε-2 atom of His-125 (3.1 Å). Besides, the O6 oxygen forms hydrogen bonds with the Oδ1 atom of Asp-141 (2.7 Å) and with the Nε-2 atom of Gln-142 (3.0 Å). Finally, in contrast to the β site, the galactose unit is also stabilized by the O3 oxygen which makes a hydrogen bond with the Nε-2 atom of His-125 (3.1 Å), and by the O2 oxygen which is at hydrogen bond distance of two water molecules (Figure 3F).

Comparison of combining sites

As indicated above, despite the β-trefoil is a highly conserved scaffold among a diverse set of proteins, their sugar-binding sites differ significantly due to the low sequence similarity exhibited by these proteins. Nonetheless, a structural analysis of the crystal structures resulting from a search at the Protein Data Bank using the keywords β-trefoil (as defined by the SCOP classification) and galactose (or lactose) clearly revealed underlying structural trends for the binding of the different galactose- or other ligands containing galactose derivatives, which were also identified within the binary complex ([LSL_{150}:(lactose)β, γ]). In particular, as shown in Table II, all combining sites included an aromatic residue (preferentially a tryptophan residue), which systematically makes van der Waals interactions with the galactose ring. This is the case even for the binding of specific ligands such as 2-deoxy-2-acetamido-β-d-galactose-4-sulfate (Liu et al. 2000) O3-sulphonyl-galactose (Liu et al. 2001) or in the particular binding modes of *N*-acetyl-D-galactosamine by the C-terminal domain of the heavy chain of the tetanus toxin (Fotinou et al. 2001) and of β-D-galactose by the mistletoe lectin 4 (Mishra et al. 2005). In addition, hydrogen bonds are created (except in the above specific cases) between the O4 oxygen of the galactose ring and an aspartic residue. Conversely, diverse ligands were identified for the galactose O3 oxygen, with also a predominance of aspartic acid residues (among them the one involved in H-bonds with the O4 oxygen).

In this context, there are two interesting aspects of LSL_{150}. First, whereas both the β- and γ-site of LSL_{150} exhibited the above general features, the α-site lacks an aromatic residue. This can be considered a sufficient condition for explaining the absence of sugar binding, since this site contains Asp-45, the putative ligand for the galactose O4 oxygen, Gln-46, the putative ligand for the galactose O6 oxygen, and also Arg-27 (equivalent to Arg-75 of the β-site and Arg-123 of the γ-site; see above). Second, the presence of a His residue interacting with the O3 oxygen of the galactose ring, a distinct feature of the γ-site of LSL_{150}, is also observed in three complexes: the ricin B-chain bound to β-D-galactose (Rutenber and Robertus 1991) the xylanase from *Streptomyces olivaceoviridis* E-86 complexed with lactose (Fujimoto et al. 2002) and in the earthworm R-type lectin c-half in complex with lactose (Suzuki et al. 2009).

Conformational changes upon lactose binding and reorganization of water molecules

The high-resolution crystal structures of unliganded and liganded states of LSL_{150} offer for the first time the opportunity to evaluate conformational changes associated with sugar binding. As expected for a lectin, lactose binding by LSL_{150} does not involve large-scale conformational changes in the polypeptide backbone (structural superposition between liganded and unliganded states provides an rmsd value of 1.1 Å for 147 Cα atoms). Nevertheless, a detailed inspection of the β and γ sites reveals a complex scenario with distinct conformational responses being observed for both sites upon lactose binding: whereas the β site can be considered mainly as preformed, a situation commonly observed in lectins (Weis and Drickamer 1996), significant changes are observed within the γ site both in the side chain conformations and in the water organization. Thus, the conformation of the lactose-interacting side chains at the β site (Arg-75, Tyr-91, Asp-93 and Asn-94) remains essentially unaffected upon lactose binding (Figure 4A). In this regard, the lack of conformational changes in lectins upon sugar binding is frequently correlated with the presence of solvent molecules occupying in the unliganded state the exact positions inhabited by hydroxyl groups of the incoming sugar (Elgavish and Shaanan 1998), which can be interpreted in terms of an existing, preformed potential field for accepting the sugar hydroxyl groups. Although crystal protein packing effects may preclude a complete analysis of the solvent reorganization within the β site of LSL_{150} (see below), some partial conclusions can be derived. In particular, a cluster of five water molecules was identified within the galactose unit binding site of the unliganded state, two of them (A2183 and A2182) situated at the exact locations occupied by the 4-OH and 6-OH groups of Gal, respectively (Figure 4A). Interestingly, the atomic B-factor values for these two water molecules (A2183: 8.2 $Å^2$; A2182: 6.2 $Å^2$) are quite similar to their liganding residues and much lower than those for the other waters (A2102: 24.8 $Å^2$; A2006: 19.0 $Å^2$; A2103: 28.7 $Å^2$), indicating a lower relative mobility of the former molecules. On the other hand, no information can be obtained on the solvent structure on the glucose unit site since it is partially occupied by the N-terminal segment of a symmetry-related molecule. As described above, water molecules primarily mediate the binding of the glucose unit at the β site and, in fact, six solvent

molecules are identified in the final complex structure directly interacting with glucose hydroxyl groups.

Regarding the γ site of LSL_{150}, lactose binding proceeds through a conformational change that mainly affects the His-125 residue (Figure 4B): the structures of unliganded and liganded states of LSL_{150} suggests that lactose binding involves a small movement of the polypeptide chain affecting the His residue, together with a displacement of its side chain of ~9 Å towards the incoming lactose, towards a position in which the Nε-2 atom of the imidazole ring is at hydrogen bond distance to the galactose unit oxygens O3 and O4 (see above). In addition, a small rearrangement of the aromatic ring of Phe-139 is observed that presumably optimizes the hydrophobic stacking interactions with the galactose ring.

To analyze solvent reorganization at the γ site of LSL_{150} upon lactose binding we have operatively distinguished three clusters of water molecules in the unliganded state (Figure 5). Cluster 1 consists of three water molecules (A2215, A2216, and A2242) (Figure 5A), two of them (A2215 and A2242) to be displaced by His-125 side chain upon lactose binding, while the other one remains at the same position (A2216) in the binary complex, at hydrogen bond distance of the Nδ-1 atom of His-125 (2.8 Å), and to the backbone nitrogen (2.7 Å) and the Oγ1 (3.3 Å) atoms of Thr-124 (Figure 5B).

Cluster 2 consists of four water molecules that will be displaced by the galactose unit in the binary complex (A2214, A2230, A2241, and A2243) (Figure 5C). As also observed in the β site, two of them (A2241 and A2243) occupy the exact locations inhabited by the 4-OH and 6-OH groups of the galactose unit, respectively. They show the lowest atomic B-factor values of the cluster (A2241: 6.2 $Å^2$; A2243: 7.6 $Å^2$; A2230: 10.3 $Å^2$; A2214: 30.7 $Å^2$), which is consistent with a preformed, fine-tuned arrangement of the liganding residues for these hydroxyl groups.

Cluster 3 consists of three water molecules (A2158, A2170, and A2229), which undergo a rearrangement upon lactose binding becoming ligands of the 3-OH, 1-OH (in the case of the α anomer), and 6-OH groups of Glc (Figures 5D and 5E), respectively. Despite this behavior consistent with the idea that structural waters may be considered as an extension of the protein surface (Toone 1994), it adds a novel feature (as far as we know) to the water-mediated lectin-sugar interactions since

the rearrangement suffered by this cluster can be considered as concerted due to the exchange of protein ligands between these water molecules. Thus, whereas in the unliganded state, water A2229 is hydrogen bonded to the NH1 atom of Arg-123, to main chain O Asn-132, to water A2230 from cluster 2, and to water A2158, in the liganded state, it forms new hydrogen bonds with the glucose 3-OH group and with the protein ligands of water A2158 (NH2 atom of Arg-76 and Nδ1 atom of His-133). In turn, this last water molecule forms a hydrogen bond with the protein ligand of water A2170 (NH1 atom of Arg-76) that now creates a new hydrogen bond with the glucose 6-OH group.

Some important aspects should be stressed in the above structural analysis of water reorganization triggered by lactose binding that may limit the conclusions drawn: firstly, we are aware that the observed rearrangement of solvent molecules upon lactose binding can also be interpreted in terms of release and binding of waters; nevertheless, the proposed concerted rearrangement of water molecules is most probable since it just involves minor, subtle reorganizations of the hydrogen bond network of already bound water molecules; secondly, it is obvious that only a partial picture of solvent organization can be obtained from crystallographic structures since only the ordered solvent molecules are observed in the electron density maps; thirdly, crystal protein packing may affect the distribution of solvent molecules within the sugar-binding sites. For instance, no information of solvent organization can be drawn at the β site for the glucose unit since the N-terminal segment of a symmetry-related LSL_{150} molecule occupies it. On the other hand, we are more confident about the conclusions drawn from the γ site since only one water molecule (A2241 of cluster 2) is involved in an interaction with a symmetry-related LSL_{150} molecule; in particular, with the Oγ1 atom of Thr-103. Finally, the resolution of the crystallographic data defines the confidence of the conclusions. In our case, we had high-resolution data for both the unliganded (1.47 Å) and liganded (1.67 Å) states and all the water molecules included in the above-considered clusters were perfectly defined in the electron density map.

In this regard, it is worth to note that the previously reported structure of the complex [LSLa:(N-acetyl-lactosamine)β, γ)] (Mancheño et al., 2005) is consistent with the scenario herein proposed for LSL_{150}, with the limitations of the lower resolution of this structure (2.7 Å; PDB code:

1W3G). In particular, the four water molecules identified at the γ-site, directly interacting with the N-acetyl-lactosamine sugar (d< 3.2 Å), are also observed in the [LSL_{150}:(lactose)β, γ)] binary complex (equivalent to waters A2250, A2251, A2256 and A2258; see Figures 3D and 3F), which provides additional supporting evidence for their direct participation in sugar binding, even more considering the remarkably different crystallization conditions for both complexes. In contrast to this, no water molecules are identified at the β-site. Nonetheless, it is interesting to see that two hydroxyl groups of the glycerol molecule identified in this site within the [LSLa:(N-acetyl-lactosamine)γ)] complex (PDB code 1W3F) occupy the locations of the water molecules A2183 and A2182 identified in LSL_{150} which in turn are situated at the exact locations occupied by the 4-OH and 6-OH groups of Gal, respectively (Figure 4A).

The correlation of binding data with structural information should always be done with caution; nevertheless, the present study permits us to draw a few qualitative conclusions. Firstly, our structural data reveals unambiguously the existence of two distinct and operative sugar-binding sites in LSL_{150}, also confirmed by the excellent fit of the ITC results to a model of two independent sites with highly different affinities so that we can define a relative low-affinity site (K_{d2} ~ 11 mM) and a high-affinity site (K_{d1} = 170 μM). In this regard, the crystal structure of the binary complex reveals that no residue from either site is *prima facie* affected by the occupancy of the other site. Besides, comparison of the β and γ sites points to the latter as the higher affinity one since both the number of direct and water-mediated sugar-protein interactions are higher, which also agrees with our previous studies on LSLa where the observed binary complexes between LSLa and lactose or N-acetyl-lactosamine only have the γ site occupied (Mancheño et al. 2005).

Fusion tag properties of LSL_{150}

In this work we have considered the working hypothesis that LSL_{150} together with its sugar-recognition role may function *in vivo* as a solubility enhancer for its membrane-interacting C-terminal module. This hypothesis was based first on the autonomous folding unit character of LSL_{150} and second on the indirect *in vitro* observation that production yields of both LSL_{150} in *E. coli* BL21 (DE3)

cells (this work) and of the deletion mutant LSL_{187} in *E. coli* Nova Blue (DE3) cells is very high, which indicates that the gene coding for LSL_{150} is a highly translated one. Indeed, this last feature has been suggested to form the basis of currently available solubility tags, which fused at the N-terminal end of a target protein improve its production yield (Malhotra 2009). In this sense, it has been recently reported that some N-terminal domains of multidomain proteins function *in vitro* as potent solubility enhancers for various C-terminal heterologous proteins (Kim et al. 2007).

Furthermore, the fact that LSLa and LSL_{187} (Tateno and Goldstein 2003), along with LSL_{150} (this work) can be effectively purified in a single-step affinity chromatography procedure on plain Sepharose® 4B indicates that LSL_{150} behaves as an excellent affinity tag, which in turn suggests potential biotechnological applications for LSL_{150}. Considering these two aspects of LSL_{150}, namely solubility enhancing and affinity tag properties, we have used an indirect *in vitro* approach to test our working hypothesis: we fused LSL_{150} to the N-terminal end of various target proteins and also included between them a short linker segment (ASSS) and a tobacco etch virus (TEV) endoprotease recognition/cleavage site (ENLYFQG) for tag removal. All the fusion proteins herein considered were successfully produced as described in the Materials and Methods, and initially purified by a single-step procedure on Sepharose® 4B. Briefly, after directly loading of the corresponding cleared cell extract and exhaustive washing of the column with binding buffer, the proteins were eluted from the column in mild conditions with the elution buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 0.04% sodium azide (w/v), and 0.2 M lactose). The obtained results (Figure 6) indicated that in all cases the desired fusion protein was effectively purified, although it was accompanied by spontaneously generated tag. After TEV digestion of the concentrated protein sample, a subsequent polishing step on 16/60 HiLoad™ Superdex® 75 (or 200) (not shown) was used to resolve the target protein from the tag and potential soluble aggregates. With the exception of PLD, which massively precipitated upon TEV digestion (see below), the rest of the proteins were obtained with a crystallographic grade quality as revealed by mass spectrometry (not shown). In fact, the validity of this preliminary purification procedure has been recently demonstrated with the purification and further crystallization of the catalytic module of Cpl-7 (Silva-Martin et al. 2010), and also of inositol 1,3,4,5,6-pentakisphosphate

kinase from *Arabidopsis thaliana,* although in this case additional purification steps were needed (Baños-Sanz et al. 2010; González et al. 2010)

Considering the latter results indicating the presence of undesired tag accompanying the fusion protein, and the evident limitation of the purification procedure as it cannot be applied to target proteins with native molecular weights similar to those of his-TEV or LSL_{150}, we have designed an optimized purification procedure of the target proteins (see Materials and Methods). Briefly, after loading the cell extract onto the Sepharose® 4B column and exhaustive overnight washing at 4 ºC, the column was equilibrated in TEV buffer (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, and 0.04% (w/v) sodium azide); then his-tagged TEV was loaded onto the column. Afterwards, the column was gently shaken for ~20 h. at 4 ºC with a roller mixer and subsequently connected in series with 1ml HisTrap FF column (GE Healthcare). The results indicated that eluted proteins were pure, with the exception of PLD (Figure 6), as also revealed by a final polishing size-exclusion chromatography step on Superdex® 75 (or 200) (not shown). In agreement with the above results, no PLD eluted from the column (Figure 6), probably due to massive aggregation of the target protein upon removal from LSL_{150}, a behavior that is frequently observed with other solubility enhancers. In fact, it has been estimated that about a quarter of the proteins expressed as MBP fusions remain insoluble or aggregate upon removal of MBP (Malhotra 2009). In turn, this emphasized that the solubility of the target proteins after tag removal depends on the target protein itself (Nallamsetty and Waugh 2006). In this regard, we believe that, as a first approach, the formation of soluble fusion proteins with C-terminal partners (insoluble by themselves) that aggregate upon tag removal should be considered as a sufficient condition for classifying a protein as solubility enhancer.

As a whole, our high-resolution crystallographic studies reveal that lactose-binding by LSL_{150} is a complex process that involves changes in the conformation of specific amino acid side chains located at the sugar-binding site, but also highlights the importance of specific solvent molecules in mediating protein-sugar interactions. In particular, we have identified for the first time a concerted rearrangement of a cluster of water molecules upon lactose binding to the γ-site. In addition, we have provided *in vitro* evidences indicating that LSL_{150} acts as a solubility tag, which in turn, may indirectly support an *in vivo* role as solubility enhancer for its natural membrane-interacting module. Moreover,

the excellent properties of LSL_{150} as affinity tag on plain agarose-based matrices have permitted us to devise a straightforward, cost-effective procedure for production and further purification of recombinant proteins in *E. coli* cells.

Materials and Methods

Bacterial strains and plasmids

Generation of LSL_{150} protein expression vector ($pKLSL_{150}$) was performed by PCR amplification of the DNA fragment encoding the amino acids 1 to 150 of the N-terminal lectin module of LSLa from *L. sulphureus* using the pair of primers LSL-F/LSL-R and the vector pET43-LSLa (Tateno and Goldstein 2003) as a template. The resulting PCR fragment flanked by NcoI and Eco RI restriction enzyme cleavage sites was cloned into pET28a(+) vector (Novagen, Germany) previously digested with the same restriction enzymes.

The vector $pKLSL_{150}$ was then used as a template for the preparation of the expression plasmid pKLSLt. In pKLSLt, the 3′-end of the LSL_{150} coding sequence was altered to include in-frame a flexible linker sequence (amino acids ASSS), the tobacco etch virus (TEV) endoprotease cleavage site (amino acids ENLYFQG) and a stop codon. This was accomplished by PCR amplification using the pair of primers (LSL-F/ LSL-TEV-R) and $pKLSL_{150}$ as a template and cloning the resulting amplified fragment into the into pET28a(+) vector (Novagen, Germany) previously digested with Nco I and Eco RI restriction enzymes.

The expression vectors coding for the selected proteins fused to LSL_t at their N-terminal ends were prepared as follows: the genes encoding the enhanced green fluorescence protein from *Aequorea victoria* (Tsien 1998) (EGFP), the endolysin from the phage Cp7 from *Streptococcus pneumonie* (Garcia et al. 1990) (Cpl-7), the sphingomyelin-dependent phospholipase from *Arcanobacterium haemolyticum* (McNamara et al. 1995) (SMD), and the sphingomyelin-dependent phospholipase from *Corynebacterium pseudotuberculosis* (McNamara et al. 1995) (PLD) were amplified by PCR using appropriate primers (Supplementary data, Table SI), from previously cloned heterologous genes. The designed primers incorporated digestion sites for EcoRI and NotI (for EGFP), and for EcoRI and

HindIII (for Cpl-7, PLD, and SMD), respectively. The resulting PCR fragments were subcloned into the pKLSL$_t$ vector previously digested with adequate enzymes in each case.

All the sequences of the primers used in the cloning procedures are in Supplementary data (Table SI), and all the different constructs were confirmed by DNA sequence analysis.

Recombinant E. coli culture

The respective plasmids were used to transform *E. coli* BL21 (DE3) cells (Novagen, Germany), the transformed cells were plated on LB agar plates containing 50 μg/mL of kanamycin. After an overnight incubation at 37 ºC, a single colony was picked with a sterile pipette and used to inoculate a 100-ml Erlenmeyer flask containing 20 ml of LB medium supplemented with 50 μg/mL of kanamycin. The culture was incubated for 4-5 h (37 ºC and 250 rpm), then 10 ml of it was used to inoculate 1 l of LB medium containing 50 μg/ml kanamycin in 5-l Erlenmeyer flasks (37 ºC and 250 rpm). When the culture turbidity (OD_{600}) reached 0.6-0.8, gene expression was induced with 0.3 mM isopropyl-β-D-thiogalactoside (IPTG) and the cultures were growth at 16 ºC continued for 20 h before harvesting the cells by centrifugation at 4000 x *g* for 15 min. Cell pellets were suspended in 25 ml 20 mM Tris-HCl, pH 8.0, 100 mM NaCl and flash frozen at -80 ºC until further use.

Protein purification

Bacterial cells were disrupted with a French Press and the resultant lysate was centrifuged at 20000 rpm in a SS34 rotor for 30 min. The soluble fraction was subsequently filtered through 0.22 μm cellulose filters and used for the purification of step.

The purification of recombinant fusion proteins containing LSL$_t$ as N-terminal tag was accomplished by affinity chromatography on Sepharose® 4B at 4 ºC using a BioLogic LP chromatography system (BioRad) and an Econo Gradient Pump (BioRad). Columns were prepared in-house using glass Econo-Column® columns (2.5 x 10 cm; BioRad). Prior to sample loading, the resin was exhaustively washed with binding buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, and 0.04% (w/v) sodium azide). Cleared cell extracts were directly loaded onto the column at 2.5 ml/min, which

was then washed with binding buffer overnight at 4 ºC. Fusion proteins were then eluted with elution buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 0.04% sodium azide (w/v), and 0.2 M lactose) at 3 ml/min. Fractions containing the eluted fusion proteins were pooled and dialyzed overnight at 4 ºC against binding buffer to remove bound lactose. Control samples from *E. coli* cell extracts over-expressing no fusion protein indicated that no protein from *E. coli* cells is adsorbed onto the column in our experimental conditions. A final, polishing size-exclusion chromatography on 16/60 HiLoad™ Superdex 75 (or 200) (GE Healthcare) was carried out to resolve the fusion protein from potential soluble aggregates and from the spontaneously produced tag (see below). Removal of the tag used for the purification of target proteins was done as follows: after loading the cell extract and the overnight washing step above described, the Sepharose® 4B column containing the bound fusion protein was equilibrated with 2-3 column volumes of TEV buffer (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, and 0.04% (w/v) sodium azide) and then, his_6-tagged TEV (~1mg total mass) was immediately loaded onto the column. The columns were subsequently gently shaken for ~20 h at 4 ºC with a roller mixer. Afterwards, the Sepharose® 4B column and a 1 ml HisTrap FF column (GE Healthcare) were connected in series. Elution of the reaction mixture was done with binding buffer at 3 ml/min and the fractions containing pure target proteins were pooled and dialyzed at 4 ºC against binding buffer to remove accompanying solutes. A final polishing step was carried out as above on 16/60 HiLoad™ Superdex 75 (or 200) (GE Healthcare). Fractions were pooled and concentrated by ultrafiltration with YM-10 membranes (Amicon). Protein materials were stored at -80 ºC. Protein purity was checked by SDS-PAGE. Protein concentration was determined by UV-VIS absorbance measurements with a Nanodrop® ND-1000 spectrophotometer, using the extinction coefficients estimated with the ExPASy sever .[51] Regeneration of the Sepharose® 4B column was done by washing with five volumes of elution buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 0.04% sodium azide (w/v), and 0.2 M lactose), and then with five additional volumes of binding buffer. Finally, protein purifications carried out with Sepharose® 4B-CL, Sepharose® 6B and Sepharose® 6B-CL rendered essentially identical results (not shown).

Isothermal titration calorimetry

Binding affinity of LSL_{150} for lactose was measured at 25º C using a MCS titration calorimeter (MicroCal, LLC, Northampton, MA). Before measurements, LSL_{150} was dialyzed extensively against 20 mM Tris-HCl, pH 8.0, 100 mM NaCl and 0.04% (w/v) sodium azide. The LSL_{150} (247 µM) solution was loaded onto the calorimetric cell and titrated by adding 1 x 1 µl, plus 20 injections (2-10 µl), of 30 mM lactose stock solution. The binding isotherms were corrected for the heats of ligand dilution and were fitted by nonlinear regression analysis using the ORIGIN ITC-software. The binding constants and the enthalpy changes were directly determined from data fitting. The free energy change was calculated as $\Delta G = -RT\ln Ka$ (R = 1.986 (cal mol) K^{-1}) and the entropy change using the Gibbs equation ($\Delta G = \Delta H - T\Delta S$).

Crystallization of LSL_{150} and [LSL_{150}:(lactose)β, γ]

Initial crystallization conditions were established using the sparse-matrix sampling technique (Jancarik and Kim 1991) with the hanging drop vapour-diffusion method at 18 ºC. Drops containing equal volumes of protein and precipitant (1+1 µl) were equilibrated against 500 µl reservoir solutions. Final optimized LSL_{150} crystallization conditions were as follows: 2 µl of protein (25-30 mg/ml) in 20 mM Tris-HCl, pH 8.0, containing 100 mM NaCl and 0.04% (w/v) sodium azide, were mixed with 2 µl of 12% (w/v) PEG 4000, 100 mM Tris-HCl, pH 8.5, 150 mM sodium acetate. Crystal plates appeared in one day and grew to an average dimension of 0.1 x 0.3 x 0.3 mm^3 in 3 days.

Additionally, LSL_{150} was co-crystallized with lactose by adding the disaccharide to the same mother liquor as above at 0.2 M final concentration. Crystals of LSL_{150} complexed with lactose appeared within two to three weeks and were 0.2 x 0.3 x 0.5 mm^3 in dimensions.

Diffraction data collection and structure determination

Crystals for diffraction data collection were flash-cooled in the cryo stream in the corresponding mother liquor supplemented with 20% (v/v) glycerol. Diffraction data from unliganded LSL_{150} crystals were collected at 100 K at beamline ID23-1 of the European Synchrotron Radiation Facility (ESRF; Grenoble, France) on an ADSC Q315R detector (X-rays wavelength 0.979 Å). The crystals belong to

the space group $P2_12_12_1$, with a = 34.58 Å, b = 59.54 Å and c = 61.62 Å. Conversely, data from crystals of LSL_{150} complexed with lactose ([LSL_{150}:(lactose)β, γ]) were recorded also at 100 K at beamline ID29 of the European Synchrotron Radiation Facility (ESRF; Grenoble, France) on an ADSC Q315R detector (X-rays wavelength 0.979 Å). These crystals belong to the trigonal space group $P3_2$, with unit-cell parameters a = 62.11 Å, b = 62.11 Å and c = 37.65 Å. Near-complete data sets were collected to maximum resolutions of 1.47 Å and 1.67 Å, for LSL_{150} and [LSL_{150}:(lactose)β, γ], respectively. Data from LSL_{150} and [LSL_{150}:(lactose)β, γ] crystals were processed and scaled using MOSFLM (Leslie 1992) and SCALA from the CCP4 package (Collaborative Computational Project Number 4 1994). The structures of LSL_{150} and the lactose complex were solved by molecular replacement with the program MOLREP (Vagin and Teplyakov 1997) using the coordinates of the N-terminal ricin-B domain of LSL_a (first 150 residues; PDB code: 1W3A) from *L. sulphureus* (Mancheño et al. 2005) as the search model. Model rebuilding and iterative refinement were done with the programs O (Jones et al. 1991) and COOT (Emsley and Cowtan 2004) and REFMAC (Murshudov et al. 1997), respectively. For the latter, a single TLS parameter was used to describe the positional distribution of the molecule as a whole. After a number of cycles of restrained refinements and addition of water molecules, the structural model was finally refined to an *R*-factor of 15.7% and an *R*-free of 19.4% (17.6% and 22.9% for the lactose complex, respectively). The final LSL_{150} model contains the complete polypeptide chain with the exception of the C-terminal Asp150 (149 amino acids) plus 252 water molecules and that for the lactose complex contains 147 amino acid residues, two lactose molecules, one glycerol and 262 water molecules. The data collection and refinement statistics are shown in Table I.

Ramachandran plots prepared using Molprobity (Davis et al. 2007) show that all nonglycine residues from LSL_{150} and from the lactose complex are in allowed regions in the Ramachandran plot. Protein structures superposition calculations were done using the SuperPose web server (Maiti et al. 2004) (http://wishart.biology.ualberta.ca/superpose). PISA server from the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) (Krissinel and Henrick 2007) was used to calculate values of interface areas. Ribbon plots were prepared using PyMOL (DeLano 2008).

Protein Data Bank accession codes

Atomic coordinates and structure factors have been submitted to the Protein Data Bank (accession codes 2Y9F and 2Y9G, for LSL_{150} and the ([LSL_{150}:(lactose)β, γ] binary lactose complex, respectively).

Funding

Financial support from the Ministerio de Ciencia e Innovación (BFU2010-17929/BMC; BFU2009-10052) and the Factoría de Cristalización (Consolider-Ingenio-2007), and CIBER de Enfermedades Respiratorias (CIBERES), an initiative of Instituto de Salud Carlos III, is greatly appreciated.

Conflict of interest statement

None declared.

Acknowledgements

We thank the ESRF (Grenoble, France) for provision of synchrotron radiation facilities. J.M.M. thanks Dr. Lourdes Infantes for her helpful insights on the interpretation of reorganization of water molecules upon lactose binding.

Abbreviations

LSL_{150}, lectin module of LSLa recombinantly produced in *E. coli*; LSL_t, LSL_{150} with a linker and TEV cleavage site at its C-terminal end; MBP, maltose binding protein; Nt-LSLa, N-terminal lectin domain of LSLa; NusA, N-utilization substance; TEV, tobacco etch virus

References

Akiba T, Higuchi K, Mizuki E, Ekino K, Shin T, Ohba M, Kanai R, Harata K. 2006. Nontoxic crystal protein from Bacillus thuringiensis demonstrates a remarkable structural similarity to beta-pore-forming toxins. *Proteins*. 63:243-248.

Baños-Sanz JI, Villate M, Sanz-Aparicio J, Brearley CA, González B. 2010. Crystallization and preliminary X-ray diffraction analysis of inositol 1,3,4,5,6-pentakisphosphate kinase from Arabidopsis thaliana. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 66:102-106.

Chervenak MC, Toone EJ. 1994. A direct measure of the contribution of solvent reorganization to the enthalpy of ligand-binding. *J Am Chem Soc*. 116:10533-10539.

Cole AR, Gibert M, Popoff M, Moss DS, Titball RW, Basak AK. 2004. Clostridium perfringens epsilon-toxin shows structural similarity to the pore-forming toxin aerolysin. *Nat Struct Mol Biol*. 11:797-798.

Collaborative Computational Project Number 4. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*. 50:760-763.

Dam TK, Brewer CF. 2002. Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chem Rev*. 102:387-429.

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, 3rd, Snoeyink J, Richardson JS, *et al.* 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*. 35:W375-383.

DeLano WL. 2008. The PyMOL Molecular Graphics System

(http://www.pymol.org).

Elgavish S, Shaanan B. 1998. Structures of the Erythrina corallodendron lectin and of its complexes with mono- and disaccharides. *J Mol Biol*. 277:917-932.

Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 60:2126-2132.

Fotinou C, Emsley P, Black I, Ando H, Ishida H, Kiso M, Sinha KA, Fairweather NF, Isaacs NW. 2001. The crystal structure of tetanus toxin Hc fragment complexed with a synthetic GT1b analogue suggests cross-linking between ganglioside receptors and the toxin. *J Biol Chem*. 276:32274-32281.

Fujimoto Z, Kuno A, Kaneko S, Kobayashi H, Kusakabe I, Mizuno H. 2002. Crystal structures of the sugar complexes of Streptomyces olivaceoviridis E-86 xylanase: sugar binding structure of the family 13 carbohydrate binding module. *J Mol Biol*. 316:65-78.

Garcia P, Garcia JL, Garcia E, Sanchez-Puelles JM, Lopez R. 1990. Modular organization of the lytic enzymes of Streptococcus pneumoniae and its bacteriophages. *Gene*. 86:81-88.

González B, Baños-Sanz JI, Villate M, Brearley CA, Sanz-Aparicio J. 2010. Inositol 1,3,4,5,6-pentakisphosphate 2-kinase is a distant IPK member with a singular inositide binding site for axial 2-OH recognition. *Proc Natl Acad Sci U S A*. 107:9608-9613.

Han JH, Batey S, Nickson AA, Teichmann SA, Clarke J. 2007. The folding and evolution of multidomain proteins. *Nat Rev Mol Cell Biol*. 8:319-330.

Hazes B. 1996. The (QxW)3 domain: a flexible lectin scaffold. *Protein Sci*. 5:1490-1501.

Hirabayashi J, Dutta SK, Kasai K. 1998. Novel galactose-binding proteins in Annelida. Characterization of 29-kDa tandem repeat-type lectins from the earthworm Lumbricus terrestris. *J Biol Chem*. 273:14450-14460.

Holm L, Sander C. 1993. Protein structure comparison by alignment of distance matrices. *J Mol Biol*. 233:123-138.

Inoue K, Sobhany M, Transue TR, Oguma K, Pedersen LC, Negishi M. 2003. Structural analysis by X-ray crystallography and calorimetry of a haemagglutinin component (HA1) of the progenitor toxin from Clostridium botulinum. *Microbiology*. 149:3361-3370.

Jancarik J, Kim SH. 1991. Sparse-matrix sampling - a screening method for crystallization of proteins. *J Appl Crystallogr*. 24:409-411.

Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A*. 47 (Pt 2):110-119.

Kim CW, Han KS, Ryu KS, Kim BH, Kim KH, Choi SI, Seong BL. 2007. N-terminal domains of native multidomain proteins have the potential to assist de novo folding of their downstream domains in vivo by acting as solubility enhancers. *Protein Sci*. 16:635-643.

Krissinel E, Henrick K. 2007. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 372:774-797.

Leslie AGW. 1992. Recent changes to the MOSFLM package for processing film and plate data. . *Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography*. 26.

Liu Y, Chirino AJ, Misulovin Z, Leteux C, Feizi T, Nussenzweig MC, Bjorkman PJ. 2000. Crystal structure of the cysteine-rich domain of mannose receptor complexed with a sulfated carbohydrate ligand. *J Exp Med*. 191:1105-1116.

Liu Y, Misulovin Z, Bjorkman PJ. 2001. The molecular mechanism of sulfated carbohydrate recognition by the cysteine-rich domain of mannose receptor. *J Mol Biol*. 305:481-490.

Loris R. 2002. Principles of structures of animal and plant lectins. *Biochim Biophys Acta*. 1572:198-208.

Maiti R, Van Domselaar GH, Zhang H, Wishart DS. 2004. SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res*. 32:W590-594.

Malhotra A. 2009. Tagging for protein expression. *Methods Enzymol*. 463:239-258.

Mancheño JM, Tateno H, Goldstein IJ, Martinez-Ripoll M, Hermoso JA. 2005. Structural analysis of the Laetiporus sulphureus hemolytic pore-forming lectin in complex with sugars. *J Biol Chem*. 280:17251-17259.

Mancheño JM, Tateno H, Sher D, Goldstein IJ. 2010. Laetiporus sulphureus lectin and aerolysin protein family. *Adv Exp Med Biol*. 677:67-80.

McNamara PJ, Cuevas WA, Songer JG. 1995. Toxic phospholipases D of Corynebacterium pseudotuberculosis, C ulcerans and Arcanobacterium haemolyticum - cloning and sequence homology. *Gene*. 156:113-118.

Melton JA, Parker MW, Rossjohn J, Buckley JT, Tweten RK. 2004. The identification and structure of the membrane-spanning domain of the Clostridium septicum alpha toxin. *J Biol Chem*. 279:14315-14322.

Mishra V, Bilgrami S, Sharma RS, Kaur P, Yadav S, Krauspenhaar R, Betzel C, Voelter W, Babu CR, Singh TP. 2005. Crystal structure of himalayan mistletoe ribosome-inactivating protein reveals the presence of a natural inhibitor and a new functionally active sugar-binding site. *J Biol Chem*. 280:20712-20721.

Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. 53:240-255.

Murzin AG, Lesk AM, Chothia C. 1992. beta-Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1 beta and 1 alpha and fibroblast growth factors. *J Mol Biol*. 223:531-543.

Nallamsetty S, Waugh DS. 2006. Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr Purif*. 45:175-182.

Notenboom V, Boraston AB, Williams SJ, Kilburn DG, Rose DR. 2002. High-resolution crystal structures of the lectin-like xylan binding domain from Streptomyces lividans xylanase 10A with bound substrates reveal a novel mode of xylan binding. *Biochemistry*. 41:4246-4254.

Nurisso A, Blanchard B, Audfray A, Rydner L, Oscarson S, Varrot A, Imberty A. 2010. Role of water molecules in structure and energetics of Pseudomonas aeruginosa lectin I interacting with disaccharides. *J Biol Chem*. 285:20316-20327.

Parker MW, Buckley JT, Postma JP, Tucker AD, Leonard K, Pattus F, Tsernoglou D. 1994. Structure of the Aeromonas toxin proaerolysin in its water-soluble and membrane-channel states. *Nature*. 367:292-295.

Rini JM, Hardman KD, Einspahr H, Suddath FL, Carver JP. 1993. X-ray crystal structure of a pea lectin-trimannoside complex at 2.6 A resolution. *J Biol Chem*. 268:10126-10132.

Rutenber E, Ready M, Robertus JD. 1987. Structure and evolution of ricin B chain. *Nature*. 326:624-626.

Rutenber E, Robertus JD. 1991. Structure of ricin B-chain at 2.5 A resolution. *Proteins*. 10:260-269.

Silva-Martin N, Molina R, Angulo I, Mancheno JM, Garcia P, Hermoso JA. 2010. Crystallization and preliminary crystallographic analysis of the catalytic module of endolysin from Cp-7, a phage infecting Streptococcus pneumoniae. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 66:670-673.

Song L, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE. 1996. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science*. 274:1859-1866.

Suzuki R, Kuno A, Hasegawa T, Hirabayashi J, Kasai KI, Momma M, Fujimoto Z. 2009. Sugar-complex structures of the C-half domain of the galactose-binding lectin EW29 from the earthworm Lumbricus terrestris. *Acta Crystallogr D Biol Crystallogr*. 65:49-57.

Svensson C, Teneberg S, Nilsson CL, Kjellberg A, Schwarz FP, Sharon N, Krengel U. 2002. High-resolution crystal structures of Erythrina cristagalli lectin in complex with lactose and 2'-alpha-L-fucosyllactose and correlation with thermodynamic binding data. *J Mol Biol*. 321:69-83.

Tahirov TH, Lu TH, Liaw YC, Chen YL, Lin JY. 1995. Crystal structure of abrin-a at 2.14 A. *J Mol Biol*. 250:354-367.

Tateno H, Goldstein IJ. 2003. Molecular cloning, expression, and characterization of novel hemolytic lectins from the mushroom Laetiporus sulphureus, which show homology to bacterial toxins. *J Biol Chem*. 278:40455-40463.

Toone EJ. 1994. Structure and energetics of protein carbohydrate complexes. *Curr Opin Struct Biol*. 4:719-728.

Tsien RY. 1998. The green fluorescent protein. *Annu Rev Biochem*. 67:509-544.

Uchida T, Yamasaki T, Eto S, Sugawara H, Kurisu G, Nakagawa A, Kusunoki M, Hatakeyama T. 2004. Crystal structure of the hemolytic lectin CEL-III isolated from the marine invertebrate Cucumaria echinata: implications of domain structure for its membrane pore-formation mechanism. *J Biol Chem*. 279:37133-37141.

Vagin A, Teplyakov A. 1997. MOLREP: an automated program for molecular replacement. *J Appl Crystallogr*. 30:1022-1025.

Vyas MN, Vyas NK, Quiocho FA. 1994. Crystallographic analysis of the epimeric and anomeric specificity of the periplasmic transport/chemosensory protein receptor for D-glucose and D-galactose. *Biochemistry*. 33:4762-4768.

Weis WI, Drickamer K. 1996. Structural basis of lectin-carbohydrate recognition. *Annu Rev Biochem*. 65:441-473.

Legends to figures

Fig. 1. Overall view of the structure of LSL_{150}. *Ribbon model* of the high resolution crystal structure of LSL_{150}, viewed in two orientations rotated by 90º. The locations of the three potential sugar binding-sites are indicated. The β-strands forming the six-stranded barrel are colored in green; those forming the hairpin triplet are in yellow, helical turns are in blue, and loops are colored in orange. The figure was prepared with PyMOL (DeLano 2008).

Fig. 2. ITC analysis of lactose binding by LSL_{150} in solution. (A) Typical ITC experiment carried out by adding aliquots of 30 mM lactose into a solution containing 247 μM LSL_{150} in 20 mM Tris-HCl, pH 8.0, 100 mM NaCl and 0.04% (w/v) sodium azide buffer at 25º C; (B) Fit of the binding isotherm to a 1:2 (protein:sugar) binding model with independent binding sites.

Fig. 3. Lactose-binding at sites β and γ of LSL_{150}. (A) Close-up stereo views of the lactose molecules bound at site β and (B) at site γ showing the 1.67 Å resolution $2|F_o|-|F_c|$ electron density map (in blue) contoured at 1σ. Close-up views of the (C) β and (D) γ carbohydrate-binding sites showing the direct and water-mediated interactions between the glucose ring of the bound lactoses and LSL_{150}. Close-up views of the (E) β and (F) γ carbohydrate-binding sites showing the direct and water-mediated interactions between the galactose ring of the bound lactoses and LSL_{150}. Water molecules are depicted as cyan spheres. Side chains involved in the interactions are labeled. Hydrogen bonds are represented with dotted lines. Distances between of potential H-bonds are also indicated.

Fig. 4. Structural comparison between unliganded and lactose-bound states of LSL_{150}. (A) Close-up view of the galactose unit-binding pocket at the β site of LSL_{150}. Water molecules identified in the unliganded state are depicted as cyan spheres. The 1.47 Å resolution $2|F_o|-|F_c|$ electron density map (in blue) contoured at 1σ is also shown for these latter molecules. Hydrogen bonds are represented as dotted lines. (B) Close-up view of the galactose unit-binding pocket at the γ site of LSL_{150}. The

structural comparison reveals that lactose binding involves a movement of the His-125 side chain of around 9 Å towards the incoming lactose. Side chains of residues that interact directly with the galactose unit of the sugar are shown as stick models (color code: orange is for the unliganded state; grey: lactose-bound state).

Fig. 5. Distribution of water molecules in the γ site of LSL_{150} in the unliganded and liganded states. (A) Cluster 1 of water molecules identified in the unliganded state of LSL_{150} in the close proximity of the side chain of His-125. (B) Solvent distribution in the close proximity of the side chain of His-125 in the lactose-bound state of LSL_{150}. A2216* indicates the equivalent water molecule to A2216 found in the bound state of LSL_{150}; this molecule corresponds to water A2206 in the 2Y9G PDB file. (C) Cluster 2 of water molecules. (D) Cluster 3 of water molecules, consisting of waters A2158, A2170, and A2229, which are distributed around the glucose unit pocket. Water A2230 is shown as a green sphere in contrast to the cyan spheres of cluster 3 waters to make clear that it belongs to cluster 2. (E) Distribution of water molecules in the same region as Figure 5D but in the lactose-bound state of LSL_{150}. The water molecules A2229*, A2158*, and A2170* correspond to A2251, A2261, and A2247 in the 2Y9G PDB file. The 1.47 Å resolution $2|F_o|-|F_c|$ electron density map (in blue) contoured at 1σ is shown for all the water molecules. Hydrogen bonds are represented as dotted lines, and distances are indicated in Å. Color code for side chains: grey, bound for of LSL_{150}; orange, unliganded LSL_{150}.

Fig. 6. SDS-PAGE analysis of the production and purification of recombinant, target proteins with LSLt as fusion tag. All the LSLt-tagged fusion proteins were purified on a one-step affinity procedure on Sepharose® 4B as described in the text. Lanes 1, 4, 7, and 10 corresponded to the samples obtained from the pooled fractions of LSLt-Cpl-7, LSLt-EGFP, LSLt-SMD, and LSLt-PLD, respectively. Fusion proteins were eluted from the Sepharose® 4B column by washing the column with Tris buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 0.04% sodium azide (w/v)) containing 0.2 M lactose. Lanes 2, 5, 8, and 11 corresponded to the target proteins Cpl-7, EGFP, SMD, and PLD, respectively, after in-column cleavage of the corresponding fusion proteins with his-TEV, and further washing of the

columns with Tris buffer. Lanes 3, 6, 9, and 12 corresponded to the samples eluted from the Sepharose® 4B column with Tris buffer containing 0.2 M lactose.

Table I. Diffraction data and refinement statistics

Values for the outermost shell are given in parentheses; rmsd, root mean square deviation

(a) Data collection and processing

Parameter	**LSL_{150}**	**[LSL_{150}:(lactose)$_{\beta,\gamma}$]**
Space group	*P* 2_1 2_1 2_1	*P* 3_2
Molecules in asymmetric unit	1	1
Unit cell (Å)	34.58, 59.54, 61.62	62.15, 62.15, 37.65
Wavelength (Å)	0.979	0.979
Resolution (Å)	34 - 1.47	53 – 1.67
Total reflections	140792	112070
Unique reflections	22387	18678
Highest resolution shell	1.55 – 1.47	1.76 – 1.67
Completeness	99.5 (98.4)	99.0 (98.1)
R_{merge} (%)	6.4 (29.7)	9.1 (32.7)
Mean I/σ(I)	16.9 (5.4)	12.6 (5.2)
Wilson B-factor (Å^2)	11.5	12.8

(b) Refinement

Parameter	**LSL_{150}**	**[LSL_{150}:(lactose)$_{\beta,\gamma}$]**
Reflections working set	21152	17696
Reflections test set	1138	961
R_{cryst}/R_{free}	0.157/0.194	0.176/0.229
Rmsd bonds (Å)	0.030	0.026
Rmsd angles (º)	2.315	2.144
Average B-factor (Å^2)	8.65	5.31 (β: 10.87; γ: 12.37)*
Number of atoms		
Protein	1212	1197
Glycerol	0	6
Lactose	0	46
Water	252	262
Ramachandran Plot		
Most favored (%)	96.6	98.6
Allowed (%)	3.4	1.4
Disallowed (%)	0	0

*Average B-factor for bound lactoses

Table II. Comparison of the combining sites present is β-trefoils. Only the main residues that interact with the galactose residue of the ligands are analysed.

PDB Code	Ligand	Aromatic residue	O4-ligands	O3-ligands
1DQO	2-deoxy-2-acetamido-β-d-galactose-4-sulfate*	Trp-117	Asn-102	
1FV3	N-acetyl-D-galactosamine*	Trp-1289	His-1271	
1FWV	O3-sulphonyl-galactose*	Trp-117		
1HWO	β-D-galactose	Trp-39	Asp-24	Asp-24, Asn-46
1IT0	β-D-galactose	Tyr-340	Asp-325	Asp-325; His-343; Asn-347
	β-D-galactose	Tyr-423	Asp-408	Asp-408; Asn-430
1KNM	β-lactose	Trp-34	Asp-19	Asp-19; His-37; Asn-41
	β-lactose	Tyr-117	Asp-102	Asp-102; Asn-124
1PUU	β-D-galactose	Trp-38	Asp-23	Lys-41
	β-D-galactose	Tyr-249	Asp-235	Asn-256
1RZO	β-D-galactose	Trp-2037	Asp-2022	Lys-2040
	β-D-galactose	Trp-2093		
1YF8	β-D-galactose*	Trp-34		Asp-19
	β-D-galactose*	Phe-75	Val-81	
	β-D-galactose*	Tyr-241		Asp-229
2AAI	β-D-galactose	Trp-37	Asp-22	Asp-22; Asn-46
	β-D-galactose	Tyr-248	Asp-234	Asp-234; His-251
2IHO	β-D-galactose	Trp-35	Asp-20	
	β-D-galactose	Trp-87	Asp-72	
2Z48	N-acetyl-D-galactosamine	Tyr-134	Asp-121	Gln-137
	N-acetyl-D-galactosamine	Tyr-181	Asp-168	Glu-184
	N-acetyl-D-galactosamine	Tyr-222	Asp-209	
	N-acetyl-2-deoxy-2-amine-galactose	Tyr-36	Asp-23	Asp-39
	N-acetyl-2-deoxy-2-amine-galactose	Trp-269	Asp-256	Asp-272
2ZQN	β-D-galactose	Trp-161	Asp-146	Asp-146; Lys-164; Asn-171
	β-D-galactose	Trp-245	Asp-230	Asp-230; Asn-252; His-248

Ligands marked with an asterisk indicate unusual, specific ligands (see the text).

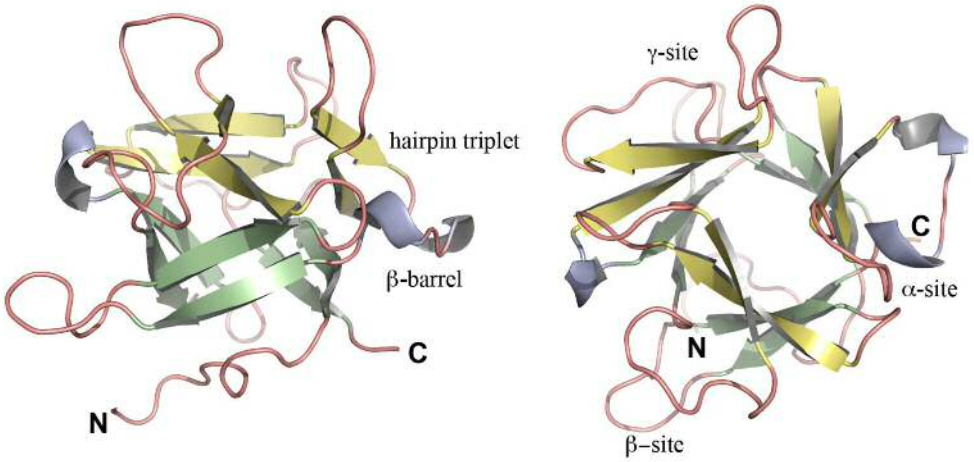

Fig. 1. Overall view of the structure of LSL150. Ribbon model of the high resolution crystal structure of LSL150, viewed in two orientations rotated by 90º. The locations of the three potential sugar binding-sites are indicated. The β-strands forming the six-stranded barrel are colored in green; those forming the hairpin triplet are in yellow, helical turns are in blue, and loops are colored in orange. The figure was prepared with PyMOL (DeLano 2008).
147x78mm (300 x 300 DPI)

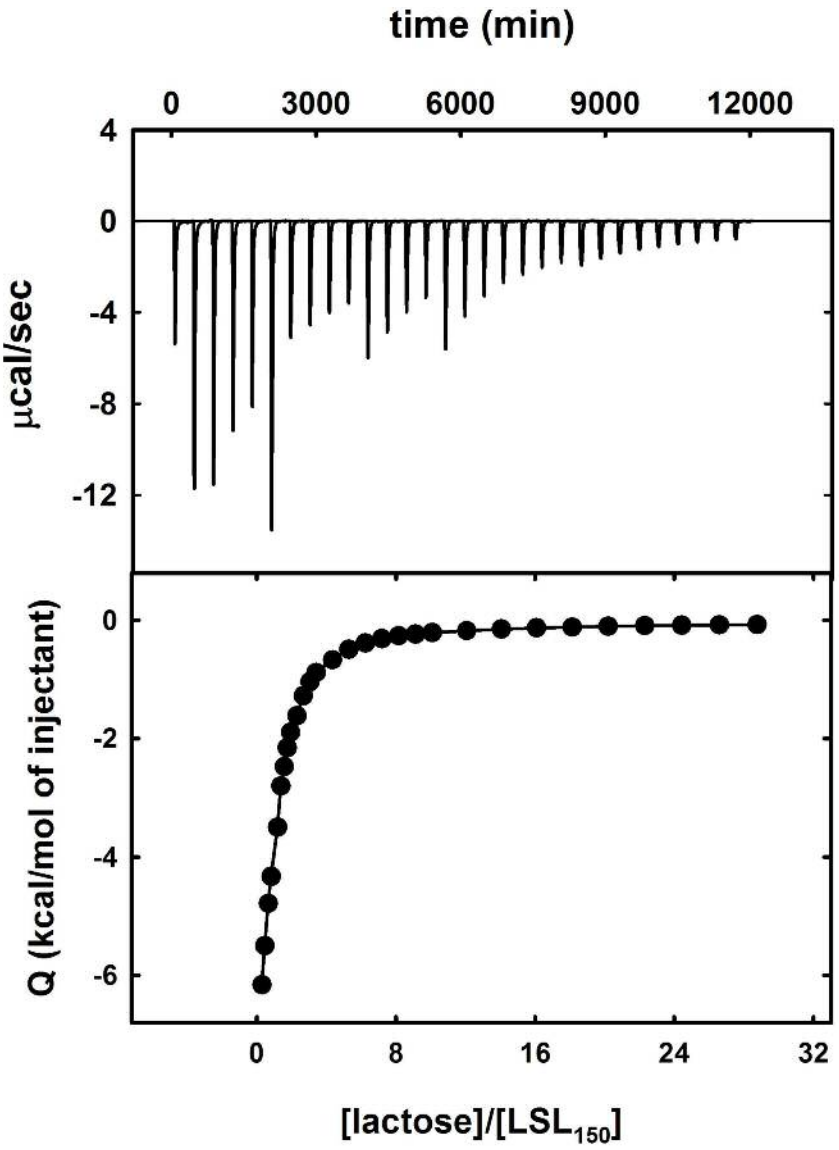

Fig. 2. ITC analysis of lactose binding by LSL150 in solution. (A) Typical ITC experiment carried out by adding aliquots of 30 mM lactose into a solution containing 247 µM LSL150 in 20 mM Tris-HCl, pH 8.0, 100 mM NaCl and 0.04% (w/v) sodium azide buffer at 25º C; (B) Fit of the binding isotherm to a 1:2 (protein:sugar) binding model with independent binding sites.

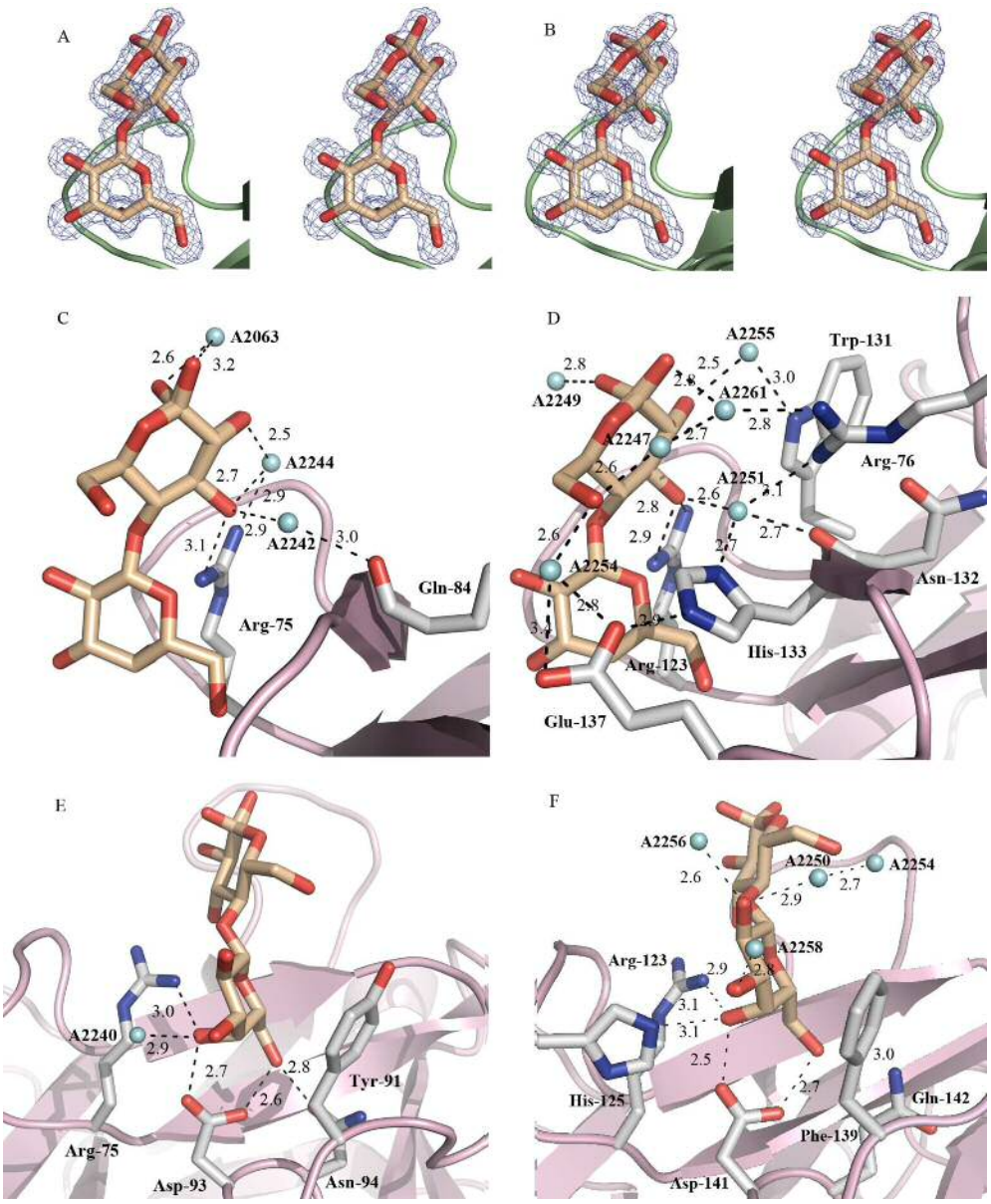

Fig. 3. Lactose-binding at sites β and γ of LSL150. (A) Close-up stereo views of the lactose molecules bound at site β and (B) at site γ showing the 1.67 Å resolution 2|Fo |- |Fc| electron density map (in blue) contoured at 1σ. Close-up views of the (C) β and (D) γ carbohydrate-binding sites showing the direct and water-mediated interactions between the glucose ring of the bound lactoses and LSL150. Close-up views of the (E) β and (F) γ carbohydrate-binding sites showing the direct and water-mediated interactions between the galactose ring of the bound lactoses and LSL150. Water molecules are depicted as cyan spheres. Side chains involved in the interactions are labeled. Hydrogen bonds are represented with dotted lines. Distances between of potential H-bonds are also indicated.
153x186mm (300 x 300 DPI)

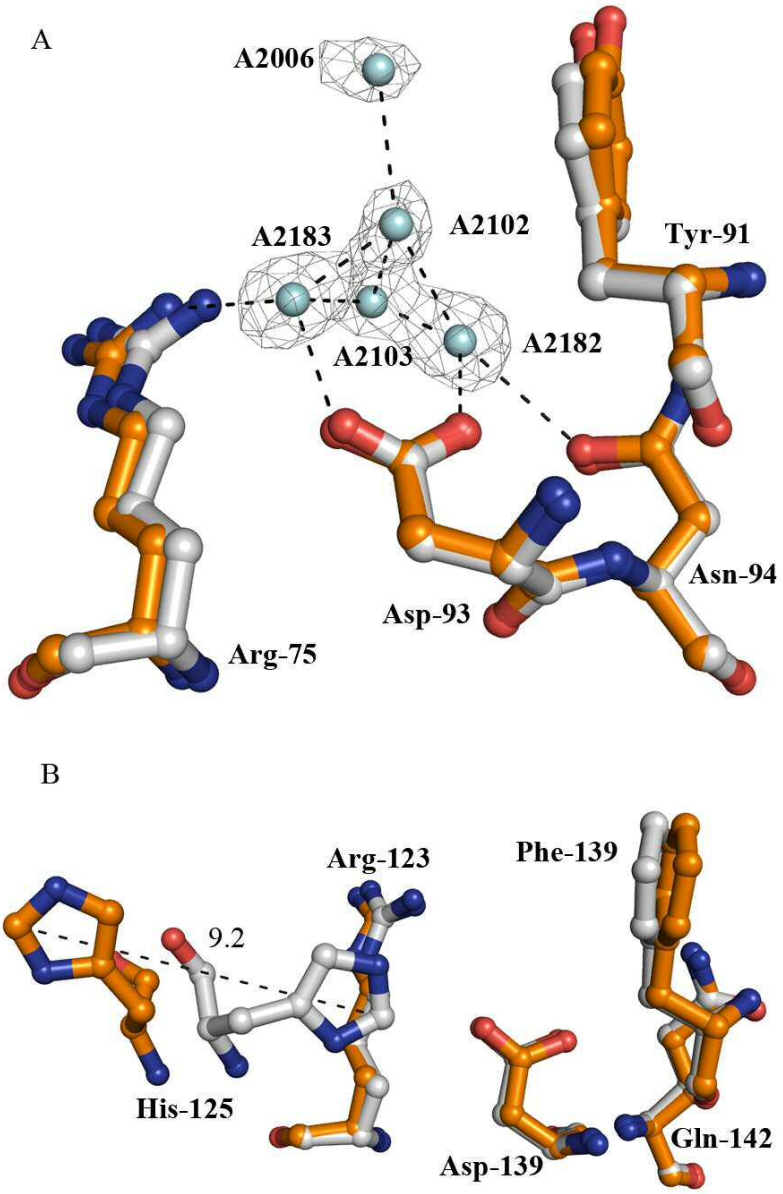

Fig. 4. Structural comparison between unliganded and lactose-bound states of LSL150. (A) Close-up view of the galactose unit-binding pocket at the β site of LSL150. Water molecules identified in the unliganded state are depicted as cyan spheres. The 1.47 Å resolution 2|Fo |- |Fc| electron density map (in blue) contoured at 1σ is also shown for these latter molecules. Hydrogen bonds are represented as dotted lines. (B) Close-up view of the galactose unit-binding pocket at the γ site of LSL150. The structural comparison reveals that lactose binding involves a movement of the His-125 side chain of around 9 Å towards the incoming lactose. Side chains of residues that interact directly with the galactose unit of the sugar are shown as stick models (color code: orange is for the unliganded state; grey: lactose-bound state).
74x113mm (300 x 300 DPI)

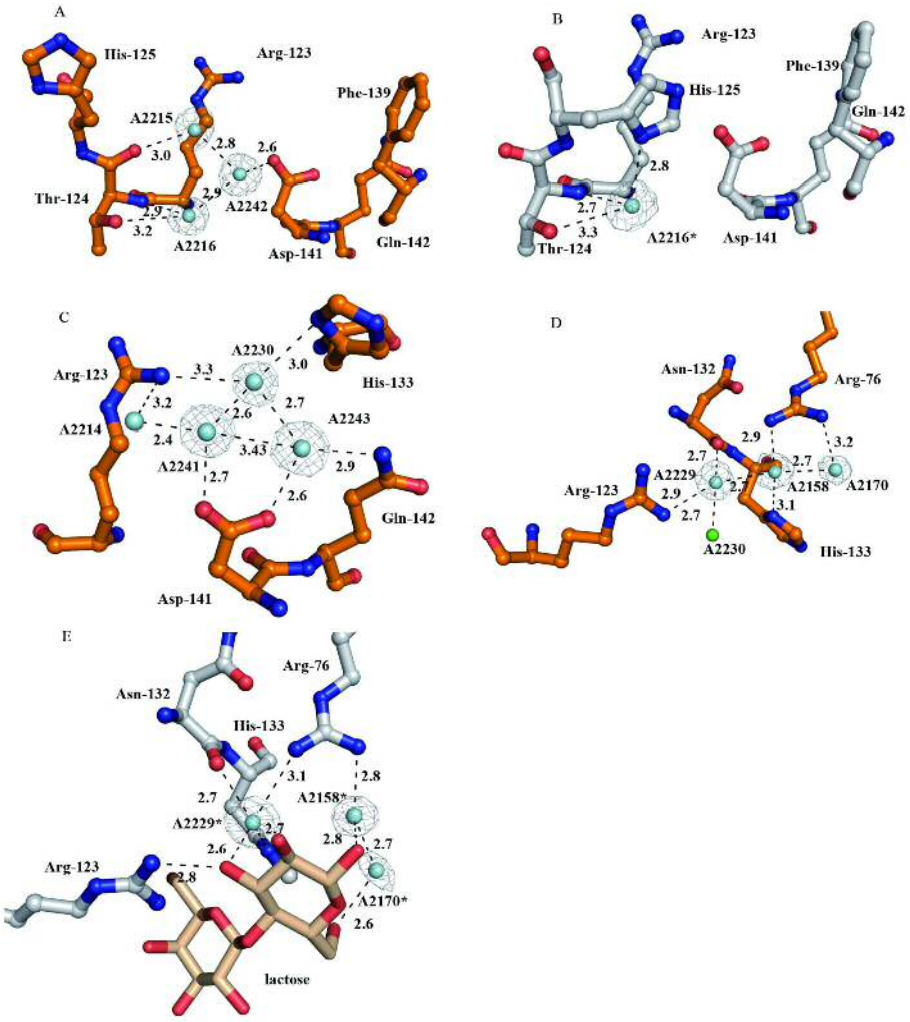

Fig. 5. Distribution of water molecules in the γ site of LSL150 in the unliganded and liganded states. (A) Cluster 1 of water molecules identified in the unliganded state of LSL150 in the close proximity of the side chain of His-125. (B) Solvent distribution in the close proximity of the side chain of His-125 in the lactose-bound state of LSL150. A2216* indicates the equivalent water molecule to A2216 found in the bound state of LSL150; this molecule corresponds to water A2206 in the 2Y9G PDB file. (C) Cluster 2 of water molecules. (D) Cluster 3 of water molecules, consisting of waters A2158, A2170, and A2229, which are distributed around the glucose unit pocket. Water A2230 is shown as a green sphere in contrast to the cyan spheres of cluster 3 waters to make clear that it belongs to cluster 2. (E) Distribution of water molecules in the same region as Figure 5D but in the lactose-bound state of LSL150. The water molecules A2229*, A2158*, and A2170* correspond to A2251, A2261, and A2247 in the 2Y9G PDB file. The 1.47 Å resolution 2|Fo |- |Fc| electron density map (in blue) contoured at 1σ is shown for all the water molecules. Hydrogen bonds are represented as dotted lines, and distances are indicated in Å. Color code for side chains: grey, bound for of LSL150; orange, unliganded LSL150.
180x199mm (300 x 300 DPI)

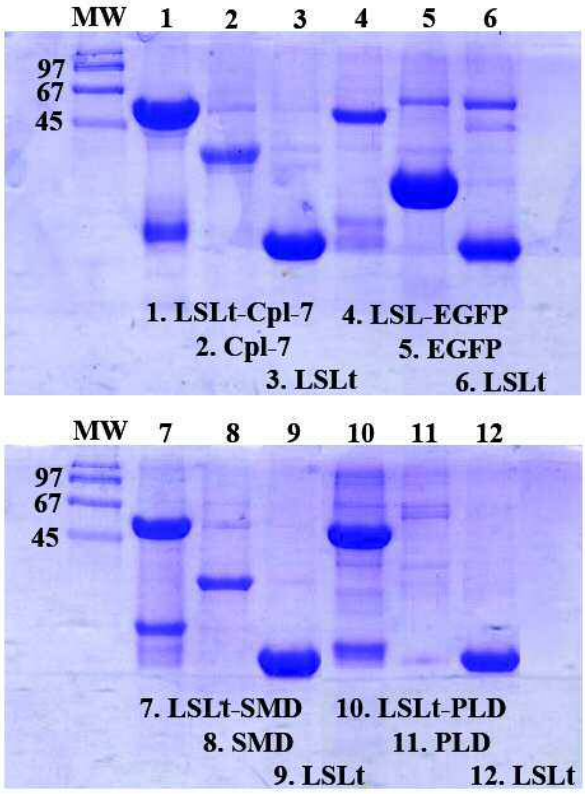

Fig. 6. SDS-PAGE analysis of the production and purification of recombinant, target proteins with LSLt as fusion tag. All the LSLt-tagged fusion proteins were purified on a one-step affinity procedure on Sepharose® 4B as described in the text. Lanes 1, 4, 7, and 10 corresponded to the samples obtained from the pooled fractions of LSLt-Cpl-7, LSLt-EGFP, LSLt-SMD, and LSLt-PLD, respectively. Fusion proteins were eluted from the Sepharose® 4B column by washing the column with Tris buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 0.04% sodium azide (w/v)) containing 0.2 M lactose. Lanes 2, 5, 8, and 11 corresponded to the target proteins Cpl-7, EGFP, SMD, and PLD, respectively, after in-column cleavage of the corresponding fusion proteins with his-TEV, and further washing of the columns with Tris buffer. Lanes 3, 6, 9, and 12 corresponded to the samples eluted from the Sepharose® 4B column with Tris buffer containing 0.2 M lactose.
99x99mm (200 x 200 DPI)